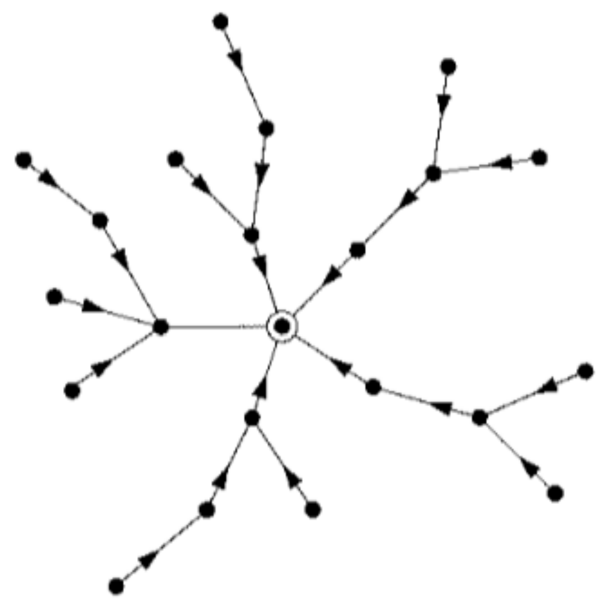
Physics 525: Homework 2

Due: Wednesday, Nov. 2'nd in class. You have the option to email me your homework sets if for some reason you cannot make it to class on that day. Having said that, there will be a strict deadline of 1PM on the due date for email submissions. No exceptions to this rule.

Policy: Feel free to discuss the problems among yourselves (in fact I encourage you to do so), however write up your own solutions separately. Please at least attempt to write something sensible for **all** questions to get full credit.

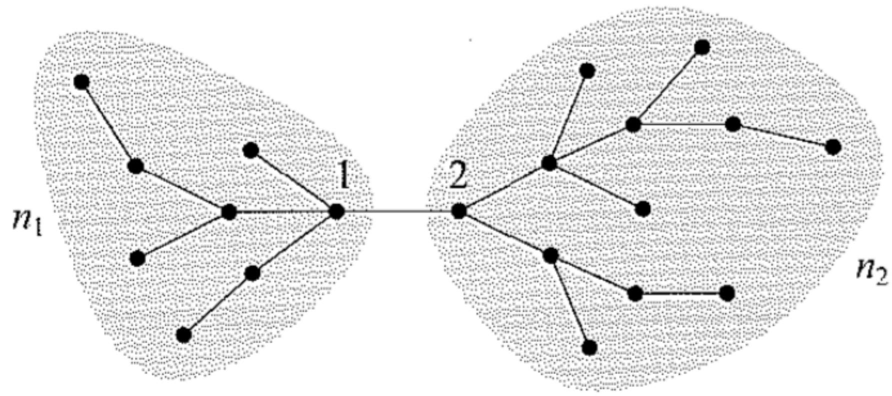
PROBLEMS

Q1. Suppose a directed network takes the form a of a tree with all edges pointing inwards towards a central vertex like so:



What is the PageRank Centrality of the *central vertex* (circumscribed by the white circle) in terms of the single parameter α (found in Eq. 7.17 in book) and the geodesic distance d_i from each vertex i to the *central vertex*?

Q2. Consider now an *undirected* tree of n vertices. A particular edge in the tree joins vertices marked **1** and **2** and divides the tree into two disjoint regions, each with n_1 and n_2 numbers of vertices as sketched below



Show that the closeness centralities C_1 and C_2 of the two vertices **1** and **2** (defined by Eq. 7.29 in book) are related by the equality:

$$\frac{1}{C_1} + \frac{n_1}{n} = \frac{1}{C_2} + \frac{n_2}{n}. \tag{1}$$

Q3. Staying with the theme of the previous question. Now consider that a particular vertex in an undirected tree of n vertices has degree k such that its removal will break up the network into k different parts each with size n_i, i.e. $n_1, n_2 \ldots n_k$. So for example imagine placing a vertex between the edges connecting **1** and **2** in the previous figure, this would then correspond to the case $k = 2$. Show that the *unnormalized* betweenness centrality x (Eq. 7.36 in book) of this central vertex of degree k is given by

$$x = n^2 - \sum_{i=1}^{k} n_i^2. \tag{2}$$

Q4. In class I showed you that the modularity of a network is given by the expression:

$$Q = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j). \tag{3}$$

While this equation is intuitively easy to understand, in practical situations it is an unwieldy expression to calculate. A more easily calculable form of the modularity can be derived in terms of the quantity

$$e_{rs} = \frac{1}{2m} \sum_{ij} A_{ij} \delta(c_i, r) \delta(c_j, s), \tag{4}$$

which is the fraction of edges that join vertices of type r to vertices of type s. (male to female for example). Along with the quantity

$$a_r = \frac{1}{2m} \sum_i k_i \delta(c_i, r), \tag{5}$$

which is the fraction of *ends of edges* attached to vertices of type r, we can rewrite Eq. (3) as

$$Q = \sum_r \left(e_{rr} - a_r^2\right). \tag{6}$$

(See Eqns 7.73 through 7.76 in book for full derivation). Now in a survey of couples in the US city of San Francisco, Catania et al [*] recorded, among other things, the ethnicity of their interviewees and calculated the fraction of couples whose members were from each possible pairing of ethnic groups. The fractions were as follows:

		Women				
		Black	Hispanic	White	Other	Total
Men	Black	0.258	0.016	0.035	0.013	0.323
	Hispanic	0.012	0.157	0.058	0.019	0.247
	White	0.013	0.023	0.306	0.035	0.377
	Other	0.005	0.007	0.024	0.016	0.053
	Total	0.289	0.204	0.423	0.084	

Assuming the couples interviewed to be a representative sample of the edges in the undirected network of relationships for the community studied, and treating the vertices as being of four types—black, Hispanic, white, and other—calculate the modularity of the network with respect to ethnicity using Eq. (6).

Q5. Suppose that a network has a degree distribution that follows the exponential form $p_k = Ce^{-\lambda k}$, where C and λ are constants and k is *discrete* (since edges are discrete):

a. Find C as a function of λ.

b. Calculate the fraction P of vertices that have degree k or greater.

c. Calculate the fraction W of ends of edges that are attached to vertices of degree k or greater.

[*]Catania, J. A, Coates, T. J. Kegels, S. and Fullilove, M. T., The population-based AMEN (AIDS in Multi-Ethnic Neighborhoods) study, *Am. J. Public Health* **82**, 284-287 (1992)

d. Hence show that for the exponential degree distribution with exponential parameter λ, the Lorenz curve—as I showed you in class for the power-law case (Eq. 8.23)—is given by

$$W = P - \frac{1 - e^{\lambda}}{\lambda} P \ln P. \tag{7}$$

e. Show that the value of W is *greater than one* for some values of P in the range $0 \leq P \leq 1$. Interpret the meaning of these “unphysical values".

Q6. Now that you are all experts in igraph the following should be a breeze. Go ahead and grab the class social network data once again from: `https://www.pas.rochester.edu/~hbarbosa/initial_network.graphml.zip` and calculate for all vertices

a. Degree Centrality

b. Closeness Centrality

c. Betweenness Centrality

d. PageRank Centrality

e. Degree Assortativity.

f. Local clustering coefficient C_i.

g. As well as the global clustering coefficient (triangles)

h. Compute now the Watts-Strogatz global clustering coefficient $C_{WS} = \frac{1}{n}\sum_i C_i$ and compare it to the standard calculation of the global clustering.

Comment upon your results. Talk about the similarities and differences you find across centralities etc. What does the degree assortativity tell you about the class? Is the clustering more or less than you expected? Basically I'm trying to see how you folks interpret the results of your data analysis.

Q7. Remember the second part of the questionnaire? This had to do with the connections you made *after* you started to attend class. That data is available at `http://www.pas.rochester.edu/~hbarbosa/final_network.graphml.zip`. Think of this as the evolution of the initial network. Repeat all the analysis of Q6 on this dataset and comment upon similarities and differences.